

## **A preference for the similar: evidence from peer mentoring for smoking cessation**

**Project summary** Although peer effects have been documented in a variety of contexts, little is known about the role played by peer characteristics in mediating those effects. Are people more likely to follow those most similar to them? If so, along what dimensions does similarity matter most? We study this question in the context of peer mentoring for smoking cessation, delivered by former smokers through a text-messaging platform. The platform serves as the basis for a randomized controlled trial to test the effectiveness and appeal of peer mentoring for smoking cessation. The experiment will include about 200 U.S. adult smokers who will be randomly assigned to a peer mentor or not. Participants in the control group will receive automated text messages used in SmokefreeTXT, a nationwide text-messaging service sponsored by the National Cancer Institute (NCI) in which users receive one to five automated messages per day for up to eight weeks. Smokers in the treatment group will receive a modified version of the same automated messages, but with additional personalized messages from a randomly assigned peer mentor. Mentors will provide responses to specific questions or comments a smoker may have after receiving the automated messages. The intervention will last 8 weeks. Our study has two aims. First, we analyze smokers' preferences for peer mentors to understand which mentor characteristics matter most to them and to investigate the role played by similarity in shaping their preferences. Second, we study whether having a peer mentor similar to themselves and/or close to their own preferences affects (i) their engagement; and (ii) their performance as measured by their quit status.

### **1. Recruitment and structure of the study**

#### **1.1 Smokers**

We will start recruiting smokers in early February 2017 and enrollment will occur on a rolling basis until the cap of 200 smokers is reached. Recruitment is expected to take 2 to 4 weeks. Smokers are recruited from the SmokefreeTXT sign-up page, targeted ads on Facebook and Google, announcements through the American Cancer Society (ACS), and referrals from others who saw the mentor advertising message. Smokers who are interested in signing up are first screened for eligibility, and then given more information about the participation

requirements and the two treatment arms.<sup>1</sup> All enrollees will be entered in two separate \$100 lottery drawings for completing a short introductory survey at the beginning of the study period and a follow-up survey about 3 months after enrollment. In between, the enrolled smokers will receive the SmokefreeTXT program with or without a peer mentor. Enrollees will be randomly assigned to either the control or the treatment group after having completed the introductory survey. Smokers assigned to the treatment group will receive a peer mentor selected at random from a set of 34 trained mentors (see below). In both arms, participants will receive 1-5 automated text messages per day for up to 8 weeks. The messages include encouragement, advice, and tips for quitting smoking. In the mentoring arm, we modified the messages to be more conversational and reworded about one automated message per day to be a conversation starter, designed to elicit a response. Upon response, the assigned mentor is available to reply and provide additional support and information. Participants in both arms will also answer 1-2 survey questions per week asking about their smoking status, current mood and craving levels. Finally, participants who report not having smoked in the last 7 days, as part of the follow-up survey, will be asked to take a saliva cotinine test. The saliva test results will be recorded through a series of photographs and shared with the study team via email or online upload; \$20 in compensation will be given for completing the test. Participants who self-report having smoked in the last 7 days will be considered to be continuing smokers and will not be required to take the test.

## **1.2 Mentors**

Mentors were recruited from October 2016 through January 2017. In total, 34 former smokers were recruited to be mentors from a variety of sources, including the Facebook pages for Smokefree.gov and the ACS Quit for Life program, and the webpage of the ACS Freshstart initiative, an online program that certifies facilitators to provide group smoking cessation counseling at work. People interested in participating first completed a screening survey.<sup>2</sup> Mentors received \$50 for completing an online training approximately two hours in length,

---

<sup>1</sup>The eligibility criteria for participants were: being at least 18 years of age, living in the U.S., being currently a smoker, having smoked > 100 cigarettes over their lifetime, not currently consuming NRT products or e-cigarettes, planning to quit smoking within 15 days, having access to a cellphone camera, webcam or tablet, and agreeing to take a saliva test.

<sup>2</sup>The eligibility criteria for mentors were: living in the US, having Internet access at least every 3 hours during the day, having smoked > 100 cigarettes over their lifetime, not currently smoking cigarettes, having quit smoking at least 6 months ago.

with modules on smoking and smoking cessation, motivational interviewing techniques and the text-messaging platform used for the study. At the end of the training, mentors were also asked to personalize a set of automated messages and to complete a mentor profile to be shared with their future mentees.<sup>3</sup> During the intervention, mentors will be assigned up to 3 mentees and will be expected to spend 5-10 minutes per day on the platform. When a new mentee is assigned, mentors will receive an introductory email with some information about the smoker’s profile (first name, time zone, gender, age, race, smoking frequency, average cigs on days smoked, years smoking, reason for smoking, reason for wanting to quit, interested in trying [quit method], quit date). Mentors will receive \$150 for mentoring and will be entered in a \$1,000 lottery drawing for which they will receive one entry plus an additional one for each mentee who is smokefree at 3 months.

### **1.3 Elicitation of smokers’ preferences for peer mentors**

In the introductory survey, participants in both intervention groups provide some information on their smoking background, their mentor preferences and basic socio-demographics. In the section on mentor preferences, participants are asked to imagine choosing a former smoker to be a peer mentor that would provide 1-on-1 support while quitting smoking. To gather insights on the mentor characteristics that matter most to them, they are asked to rank 7 selection criteria in order of importance, including 3 demographic variables (age, gender, race/ethnicity) and 4 smoking-related variables (heavy/light smoker, reason for smoking, reason for quitting, method used to quit). After having submitted their ranking, smokers are asked to pick a value for each of the 7 variables (e.g., “If you could choose your mentor’s method of quitting, what would you choose?”) by making a single selection from a list of categories that includes the option “no preference”. Table 1 provides a summary of each of the 7 variables. We note that the preferences expressed by participants have no impact on their mentor assignment, as the matching is random.

---

<sup>3</sup>Mentees receive a link to their mentor profile with information on the following: name, profile picture, gender, age, story as a former smoker, smokefree since MM/DD/YYYY, years as a smoker, average cigarettes per day before quitting, main reason for quitting, methods used to quit smoking, reason for becoming a mentor. A few other (non required) items were only completed by a subset of mentors, creating some variation in terms of what each mentee knows about his mentor from the start.

Table 1: List of individual characteristics

Individual characteristic $k$	number of categories	list of categories
Gender	3	Male, Female, No preference
Age	9	18-21, 21-29, ..., 70-79, 80+, No preference
Race/ethnicity	8	White, Black/African American, Hispanic or Latino, American Indian or Alaska Native, Asian, Native Hawaiian, or other Pacific Islander, other/multiracial, No preference
Heavy/light smoker	5	Light, non-daily (intermittent, casual), Light, daily ( $< 10$ cigs/day) Average, daily (between 10 and 20 cigs/day) Heavy, daily ( $> 20$ cigs/day), No preference
Reason for smoking	9	Enjoyment, Stress relief, Weight control Boredom relief, Aid to concentration, Aid to socializing Pain relief, Liking being a smoker, No preference
Reason for quitting	9	For family, To improve personal health, To live longer, To stop being an addict, To save money, To stop smelling like smoke, To improve general appearance, Because society does not approve of smoking, No preference
Method used to quit	9	Self-help pamphlet, brochure or other materials, Internet quitting program, In-person counseling, Telephone counseling Nicotine replacement product, Stop-smoking medication, Electronic cigarettes Counseling and medication/nicotine replacement, No preference

## 2. Data analysis

We will use the data generated from this randomized controlled trial to study two specific questions. First, we will study smokers’ preferences for peer mentors to understand what mentor characteristics are perceived as most valuable by smokers and whether participants have a preference for similarity. Second, using the subsample of smokers who will be assigned a mentor ( $N \approx 100$ ), we will study whether being assigned to a mentor closer to one’s own characteristics or closer to one’s preferred characteristics affects engagement and likelihood of quitting. The evaluation of the peer mentoring intervention, relative to the control group, will be conducted in a separate study registered on ClinicalTrials.gov (ID NCT03048786).

### 2.1 Preferences for peer mentors

For this question, we will use data from the entire sample of 200 smoker participants who provided a rank ordering and expressed a preference for each of the 7 mentor characteristics mentioned in 1.3.

**Preferred characteristics** Each smoker  $i$  will be asked to assign to each mentor characteristic  $k$  a rank number  $r_{ik} \in \{1, \dots, 7\}$ , where 1 is most preferred and 7 is least preferred. We will first study what criteria are perceived as most important by comparing the distribution of ranks  $F_k(r_k)$  for each characteristic  $k$ . We will test for differences in mean rank, first across all 7 characteristics separately and then grouping characteristics into socio-demographic and smoking-related variables. Significance will be assessed through an F-test after the regression of the rank number  $r_{ik}$  on dummies for each characteristic (standard errors clustered at the individual level). Finally, for each criterion  $k$ , we will study whether specific smoker characteristics drive preference for that particular criterion by running a linear regression of  $r_{ik}$  on a set of individual covariates  $\mathbf{X}_i$ , including basic socio-demographics and smoking-related variables.<sup>4</sup>

**Preference for similarity** We will then turn to the core analysis of preference for similarity: do smokers tend to prefer mentors more similar to themselves and for what criteria does similarity matter most? For each mentor characteristic (e.g., gender), we will first study the distribution of preferences (e.g., male/female/no preference), both unconditional and conditional on the smoker's category for that specific characteristic. Conditional distributions will be compared using  $\chi^2$  tests. We will then construct similarity indices to aggregate information over all characteristics. Let  $v_{ik}$  be smoker  $i$ 's value for characteristic  $k$  (e.g.,  $v_{ik} = 0$  for "no preference",  $=1$  for "male",  $=2$  for "female") and let  $\tilde{v}_{jk}^i$  be his preferred value for mentor  $j$  on characteristic  $k$ . A smoker has a preference for similarity on characteristic  $k$  if  $\tilde{v}_{jk}^i = v_{ik}$ . For each smoker  $i$ , we will construct the following Preference for Similarity Index:

$$PSI_i = \sum_{k=1}^7 \pi_{ik} \mathbb{1}_{\{\tilde{v}_{jk}^i = v_{ik}\}}$$

where  $\pi_{ik} > 0$  is the weight given to characteristic  $k$  in the computation of the index. We will consider two specifications for the  $\pi_{ik}$ 's, unweighted ( $\pi_{ik} = \frac{1}{7}$  for all  $i, k$ ) and weighted based on criterion ranking, where  $\pi_{ik} = \frac{8-r_{ik}}{\sum_k r_{ik}}$ . We will study the distribution of this index, the

---

<sup>4</sup>In all regressions of this study, the vector  $\mathbf{X}_i$  will contain gender, age, race/ethnicity (white, black, Hispanic, other), income, number of cigs/day, and number of past quit attempts. In our main specifications, we will only use this primary list of covariates. In a second set of specifications, the list of covariates will be expanded to include: years smoking, preferred method for quitting, reason for smoking, reason for quitting, education, marital status, number of children, immigrant vs native.

criteria that contribute to higher index values, and smoker characteristics that may predict a higher preference for similarity.<sup>5</sup> If participants make selections partly at random, the likelihood of a match will be higher for criteria with fewer categories. Therefore we will also compare the distribution of  $PSI_i$  in our dataset to the one obtained with a simulated dataset of 200 smokers who made selections at random. Finally, we will study how smokers' preference for similarity depends on their rank ordering of the 7 criteria: is similarity more or less important on criteria for which a smoker has a stronger preference? To do so, we will compare the mean rank of characteristics for which a match is preferred ( $\tilde{v}_{jk}^i = v_{ik}$ ) to those for which no match is preferred ( $\tilde{v}_{jk}^i \neq v_{ik}$ ). Significance will be assessed by an F-test after a linear regression of  $r_{ik}$  on dummies  $\mathbb{1}_{\{\tilde{v}_{jk}^i = v_{ik}\}}$  and  $\mathbb{1}_{\{\tilde{v}_{jk}^i \neq v_{ik}\}}$  (standard errors clustered at the individual level).

## 2.2 Impact of similarity on engagement and likelihood of quitting

The second part of our study will use the subsample of 100 smokers who were assigned a peer mentor in order to understand whether being assigned a mentor more similar to oneself affects engagement and likelihood of quitting.

**Definition of the outcomes variables** We will consider two sets of outcomes pertaining to smoking status and level of engagement of a smoker  $i$  who was assigned mentor  $j$ :

*Smoking status:* our primary outcome measure is a dummy variable equal to 1 if a participant passed the 3-month saliva test (i.e., tested negative for salivary cotinine); smokers who do not take the test or fail to upload the results will be considered to be continuing smokers as part of the intent-to-treat analysis. As a secondary outcome, we will also look at the change after 3 months in the number of cigarettes smoked  $\Delta cigs$  reported by participants in the follow-up survey; for non respondents, we will assume that consumption did not change.

*Level of engagement:* our primary outcome is the number of days from enrollment  $d$  since the participant last replied to a message asking about the person's status (e.g., mood and cravings). We will also study communication intensity through the total number of messages

---

<sup>5</sup>More precisely, we will run linear regressions of the form:

$$PSI_i = \beta_0 + \mathbf{D}_k \beta_1 + \mathbf{X}_i \beta_2 + \epsilon_i$$

where  $\mathbf{D}_k$  is a vector of dummy variables for whether  $\tilde{v}_{jk}^i = v_{ik}$  and  $\mathbf{X}_i$  is as in footnote 4.

$m$  sent by a participant to his mentor. As a final measure of engagement, we will look at the proportion of participants who unsubscribe from the text-messaging intervention prior to the end of the intervention period (dummy *unsubscribed*).

**Actual versus ideal mentor** For each outcome measure, we will study the effect of being assigned a mentor closer to one’s preferences and/or closer to one’s own characteristics. In order to do so, we consider two types of indices defined below.

*Preference Match:* The first index encodes the degree to which the characteristics of the assigned mentor  $j$  match the preferences expressed by a participant  $i$ . Denote as  $\mathcal{P}^i := \{k \mid \tilde{v}_{jk}^i \neq 0\}$  the set of all mentor characteristics for which smoker  $i$  expressed a specific preference (0 = “no preference”), and let  $v_{jk}$  be mentor  $j$ ’s actual value for characteristic  $k$ . We consider the following Preference Match Index:

$$PMI_{ij} = \frac{\sum_{k \in \mathcal{P}^i} \pi_{ik} \gamma(v_{jk}, \tilde{v}_{jk}^i)}{\sum_{k \in \mathcal{P}^i} \pi_{ik}}$$

where  $\gamma(v_{jk}, \tilde{v}_{jk}^i) = \mathbb{1}_{\{v_{jk} = \tilde{v}_{jk}^i\}} + 0.5 \cdot \mathbb{1}_{\{v_{jk} = \tilde{v}_{jk}^i \pm 1\}}$  for characteristics with ordinal categories (age and heavy/light smoker) and  $\gamma(v_{jk}, \tilde{v}_{jk}^i) = \mathbb{1}_{\{v_{jk} = \tilde{v}_{jk}^i\}}$  otherwise. In words,  $PMI_{ij}$  measures the extent to which the assigned mentor  $j$  matches smoker  $i$ ’s preferences when those were expressed. Intuitively, we expect participants to be more motivated and engaged if they were assigned a mentor who is closer to the preferences they expressed.

*Match Homogeneity:* The second index captures the degree of homogeneity of the match, i.e., how similar mentor and mentee are based on observables. To do so, we use the information collected during the first survey for both mentor and mentee regarding a large number of individual characteristics (n: nominal, o: ordinal, c: continuous):

Socio-demographics: gender (n), age (c), race (n), country of birth (n), marital status (n), number of kids (c), education level (o), income (o).

Smoking-related variables: number of cigs/day (c), number of years smoking (c), main reason for smoking (n), main reason for quitting (n), methods recommended/preferred (n), number of smokers in the household (c), number of quit attempts (c), importance of quitting (o),

self-efficacy (o), perceived difficulty of quitting (o).

We then construct the following Match Homogeneity Index based on the set  $\mathcal{K}$  of individual characteristics listed above:

$$MHI_{ij} = \sum_{k \in \mathcal{K}} \pi_{ik} \gamma(v_{ik}, v_{jk})$$

where  $\pi_{ik} = \frac{1}{|\mathcal{K}|}$  and  $\gamma(v_{ik}, v_{jk}) = \begin{cases} \mathbb{1}_{\{v_{ik} = v_{jk}\}} & \text{if } k \text{ is nominal} \\ \mathbb{1}_{\{v_{ik} = v_{jk}\}} + 0.5 \cdot \mathbb{1}_{\{v_{ik} = v_{jk} \pm 1\}} & \text{if } k \text{ is ordinal} \\ 1 - \left| \frac{v_{ik} - v_{jk}}{v_{max} - v_{min}} \right| & \text{if } k \text{ is continuous} \end{cases}$

where  $v_{max}$  and  $v_{min}$  are the upper and lower bound responses for the entire sample of mentors and mentees.<sup>6</sup>

Below is a summary table of the 3 different indices that will be constructed in this study.

Table 2: Summary of similarity indices

Index name	Description
Preference for Similarity Index ( <i>PSI</i> )	similarity between own attributes and preferred mentor attributes
Preference Match Index ( <i>PMI</i> )	concordance of assigned mentor attributes and preferred mentor attributes
Match Homogeneity Index ( <i>MHI</i> )	concordance of own attributes and assigned mentor attributes

**Regression specifications** For each outcome  $y_{ij} \in \{passed\ test_{ij}, \Delta cigs_{ij}, d_{ij}, m_{ij}, unsubscribed_{ij}\}$  defined above and  $Index_{ij} \in \{PMI_{ij}, MHI_{ij}\}$ , we will estimate linear regression models of the form

$$y_{ij} = \tilde{\mathbf{X}}_i \beta + \gamma Index_{ij} + \alpha_j + \epsilon_{ij}$$

where  $\alpha_j$  is a set of mentor dummies and  $\tilde{\mathbf{X}}_i$  is a vector of control variables (including a constant) that contains  $\mathbf{X}_i$  described in footnote 4 as well as the smoker's quit date, baseline cell phone usage, and number  $\tilde{m}_{ji}$  of messages received from mentor  $j$  (adjusted for script length). Standard errors will be clustered at the mentor level. We hypothesize that  $\gamma > 0$ , i.e., smokers will be more engaged and less likely to smoke if they were assigned a mentor they like/similar to them.

---

<sup>6</sup>That is,  $v_{min} = \min\{\min_{i \in \mathcal{I}} v_i, \min_{j \in \mathcal{J}} v_j\}$  and  $v_{max} = \max\{\max_{i \in \mathcal{I}} v_i, \max_{j \in \mathcal{J}} v_j\}$



We will also run regressions with both  $PMI_{ij}$  and  $MHI_{ij}$  as explanatory variables, and interacting those two indices with the  $PSI$  index. We hypothesize that the effect of similarity will be stronger for those who prefer similarity. Finally, we will ask whether the characteristics for which participants would prefer a mentor similar to themselves differ from those for which a match predicts a higher likelihood of engagement and success. In order to do so, we will consider regressions of each outcome  $y_{ij}$  on a set of dummies  $\mathbb{1}_{\{v_{jk}=v_{ik}\}}$ , in order to estimate the relative contribution of a match on characteristic  $k$  to the prediction of  $y_{ij}$ . Findings will be contrasted with the results obtained for the relative weight of each  $k$  in the  $PSI$  index (see footnote 5).